Computationally efficient objective function for algebraic codebook optimization in ACELP

Tom Bäckström

International Audio Labs Erlangen, Fraunhofer IIS, Erlangen, Germany

Abstract

Speech coding with the ACELP paradigm is based on a least squares algorithm in a perceptual domain, where the perceptual domain is specified by a filter. This article shows that the computational complexity of this conventional definition of the least squares problem can be reduced by taking into account the impact of the zero impulse response into the next frame. The proposed modification introduces a Toeplitz structure to a correlation matrix appearing in the objective function, which simplifies the structure and reduces computations. It is shown that the proposed method reduces computational complexity up to 17% without reducing perceptual quality.

Index Terms: ACELP, speech coding, Toeplitz matrix, perceptual model, computational complexity

1. Introduction

In speech coding, Code Excited Linear Prediction (CELP) with an algebraic residual codebook (ACELP) is the approach of choice in main stream codecs such as [1, 2, 3]. ACELP is based on modelling the spectral envelope by a linear predictive (LP) filter, the fundamental frequency of voiced sounds by a long time predictor (LTP) and the prediction residual by an algebraic codebook. The LTP and algebraic codebook parameters are optimized by a least squares algorithm in a perceptual domain, where the perceptual domain is specified by a filter.

The computationally most complex part of ACELP-type algorithms, the bottleneck, is optimization of the residual codebook. The only currently known optimal algorithm would be an exhaustive search of a size N^p space for every sub-frame, where at every point, an evaluation of $\mathcal{O}(N^2)$ complexity is required. Since typical values are sub-frame length $N = 64$ (i.e. 5ms) with $p = 8$ pulses, this implies more than 10^{20} operations per second. Clearly this is not a viable option. To stay within the complexity limits set by hardware requirements, codebook optimization approaches have to operate with non-optimal iterative algorithms. Many such algorithms and improvements to the optimization process have been presented in the past, for example [1, 4, 5, 6, 7].

The main purpose of this article is to demonstrate that by a slight modification of the objective function, complexity in the optimization of the residual codebook can be further reduced. This reduction in complexity comes without reduction in perceptual quality. As an alternative, since ACELP residual optimization is based on iterative search algorithms, with the presented modification, it is possible to increase the number of iterations without an increase in complexity, and in this way obtain an improved perceptual quality.

Observe that both the conventional and the modified objective functions model perception and strive to minimize perceptual distortion. However, the optimal quantization of the conventional approach is not necessarily optimal with respect to the modified objective function and vice versa. This alone does not mean that one approach would be better than the other, but analytic arguments do show that the modified objective function is more consistent. Specifically, in contrast to the conventional objective function, the proposed approach treats all samples within a sub-frame equally, with consistent and well-defined perceptual and signal models.

The proposed modification can be applied such that it changes only the optimization of the residual codebook. It does therefore not change the bit-stream structure and can be applied in a back-ward compatible manner to existing ACELP codecs.

In the following, in the interest of brevity, we shall by the word *frame* denote a temporal block where the perceptual and signal models are constant. This corresponds to the conventional 5ms unit of a sub-frame, but does not restrict the results in generality.

2. Objective Functions

ACELP is based on a signal model consisting of the LP, the LTP and an algebraic codebook, which represent, respectively, the spectral envelope, fundamental frequency and the prediction residual. In the current context, LTP does not play a significant role and to simplify notation, we will omit it. Also tools like pre-emphasis and formant-enhancement are omitted in the interest of clarity. The reader can readily employ conventional methods to re-introduce these tools [1, 2].

The codec output is evaluated in a perceptual domain, where the perceptual model is represented by a filter. Let $X(z)$, $A^{-1}(z)$, $W(z)$ be the Z-transforms of the speech signal, signal model and perceptual weighting filter, respectively, and let $Y(z) = W(z)A^{-1}(z)X(z)$ be the perceptual domain speech signal. Moreover, let the variables with hats, $\hat{X}(z)$, $\hat{Y}(z)$, denote their respective quantized values. Since the product $H(z) = W(z)A^{-1}(z)$ is also a filter, we can use this more compact notation to describe the perceptually weighted signal model, $Y(z) = H(z)X(z)$.

In time-domain we have

$$y_n = h_n * x_n = \sum_{k=0}^{K-1} h_k x_{n-k} \tag{1}$$

where y_n, h_n and x_n are the coefficients of $Y(z)$, $H(z)$ and $X(z)$, respectively. Both the conventional and the proposed objective functions try to minimize the perceptual coding error, that is, to minimize the difference between the original weighted signal y_n and the quantized and weighted synthesis signal $\hat{y}_n$. The difference between the two is in the windowing.

In the conventional approach, y_n is windowed such that when quantizing x_n on a range $N_1 \leq n < N_2$, the energy difference between y_n and $\hat{y}_n$ on the range $N_1 \leq n < N_2$ is minimized. In other words, the conventional objective function

can be written as

$$\min_{\hat{x}_{N_1}\cdots\hat{x}_{N_2-1}} \sum_{n=N_1}^{N_2-1} (y_n - \hat{y}_n)^2. \tag{2}$$

Equivalently, we can write

$$\min_{\hat{x}_{N_1}\cdots\hat{x}_{N_2-1}} \sum_{n=N_1}^{N_2-1} \left[\sum_{k=n-N_1+1}^{K-1} h_k(x_{n-k} - \hat{x}_{n-k}) + \sum_{k=0}^{n-N_1} h_k(x_{n-k} - \hat{x}_{n-k}) \right]^2. \tag{3}$$

Observe that the first summation has only terms of $\hat{x}_n$ where $n < N_1$ and they are thus, with respect to the optimization, constant.

By assuming that the frame-length is N and that h_n is also of length N, and by defining matrices $\mathbf{y}_k = [y_{Nk}, \ldots, y_{N(k+1)-1}]^T$ and $\mathbf{x}_k = [x_{Nk}, \ldots, x_{N(k+1)-1}]^T$ we have in matrix form

$$\begin{aligned} &\min_{\hat{\mathbf{x}}_k} \|\mathbf{y}_k - \hat{\mathbf{y}}_k\|^2 \\ &\quad = \min_{\mathbf{x}_k} \|\mathbf{U}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}) + \mathbf{L}(\mathbf{x}_k - \hat{\mathbf{x}}_k)\|^2 \\ &\quad = \min_{\mathbf{x}_k} \|\mathbf{y}_k - \mathbf{U}\hat{\mathbf{x}}_{k-1} - \mathbf{L}\hat{\mathbf{x}}_k\|^2 \end{aligned} \tag{4}$$

where $\mathbf{U}$ and $\mathbf{L}$ are upper and lower triangular matrices

$$\mathbf{L} = \begin{bmatrix} h_0 & 0 & \ldots & 0 \\ h_1 & h_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ h_{N-1} & h_{N-2} & \ldots & h_0 \end{bmatrix} \tag{5}$$

and

$$\mathbf{U} = \begin{bmatrix} 0 & h_{N-1} & \ldots & h_1 \\ 0 & 0 & \ddots & \vdots \\ \vdots & & \ddots & h_{N-1} \\ 0 & 0 & \ldots & 0 \end{bmatrix} \tag{6}$$

such that the full convolution matrix is

$$\mathbf{H} = \begin{bmatrix} \mathbf{L} \\ \mathbf{U} \end{bmatrix} = \begin{bmatrix} h_0 & 0 & \ldots & 0 \\ h_1 & h_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ h_{N-1} & h_{N-2} & \ldots & h_0 \\ 0 & h_{N-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & h_{N-2} \\ 0 & \ldots & 0 & h_{N-1} \end{bmatrix}. \tag{7}$$

Observe that the term $\mathbf{U}\hat{\mathbf{x}}_{k-1}$ is known as the zero-impulse response (ZIR) since it corresponds to the weighted impulse response of the signal model with the excitation $\mathbf{x}_{k-1}$ extended with zeros.

The first central observation of this work is that in Eq. 4, when minimizing for $\hat{\mathbf{x}}_k$, the formula contains $\mathbf{U}\hat{\mathbf{x}}_{k-1}$, a term depending on the quantisation of the previous frame. Since that term has been already quantised, we cannot change it anymore although it could be sub-optimal in view of the current frame. However, if we window x_n instead of y_n, such that the windowed residual x'_n has $x'_n = x_n$ in the range $N_1 \leq n < N_2$

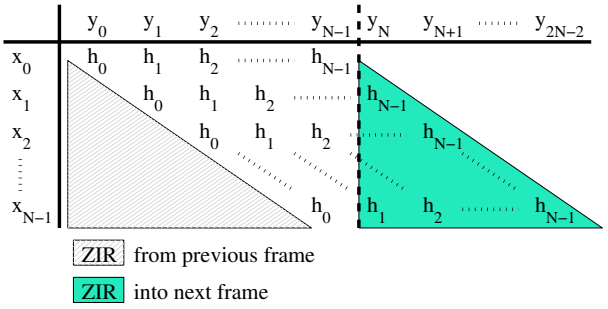

Figure 1: Illustration of the difference between the conventional and consistent objective functions, as well as of the zero-impulse responses (ZIR).

and $x'_n = 0$ elsewhere, and $y'_n = h_n * x'_n$, we can modify the optimization problem to

$$\min_{\hat{x}_{N_1}\cdots\hat{x}_{N_2-1}} \sum_{n=N_1}^{N_2+K-1} (y'_n - \hat{y}'_n)^2 \tag{8}$$

which corresponds to

$$\min_{\hat{x}_k} \|\mathbf{y}'_k - \hat{\mathbf{y}}'_k\|^2 = \min_{\hat{x}_k} \|\mathbf{H}(\mathbf{x}_k - \hat{\mathbf{x}}_k)\|^2. \tag{9}$$

Note that this form omits the dependence to the previous frame.

The second observation is that the norm in Eq. 9 has more structure than that in Eq. 4. From Eq. 4 we have

$$\|\mathbf{y}_k - \mathbf{U}\hat{\mathbf{x}}_{k-1} - \mathbf{L}\hat{\mathbf{x}}_k\|^2 = b + \mathbf{d}^T\hat{\mathbf{x}}_k + \hat{\mathbf{x}}_k^T\mathbf{C}\hat{x}_k \tag{10}$$

where b is a constant scalar, $\mathbf{d}$ a constant vector and $\mathbf{C} = \mathbf{L}^T\mathbf{L}$. In comparison, from Eq. 9,

$$\|\mathbf{H}(\mathbf{x}_k - \hat{\mathbf{x}}_k)\|^2 = \tilde{b} + \tilde{\mathbf{d}}^T\hat{\mathbf{x}} + \hat{\mathbf{x}}_k^T\mathbf{R}\hat{\mathbf{x}}_k \tag{11}$$

where $\mathbf{R} = \mathbf{H}^T\mathbf{H}$ is the autocorrelation matrix of the sequence h_n. The main result of the current work is then that, whereas $\mathbf{C}$ does not have any useful structure other than symmetry, the matrix $\mathbf{R}$ has symmetric Toeplitz structure, whereby calculation, access and storage of $\mathbf{R}$ is straightforward. It follows that evaluation of the modified objective function in Eq. 9 is simpler than the conventional approach of Eq. 4.

Concurrently, observe that if we have perfect quantisation $\mathbf{x}_k = \hat{\mathbf{x}}_k$, for all k, then $\mathbf{y}_k = \hat{\mathbf{y}}_k$ in both cases. Both objective functions thus posses the perfect-reconstruction property, whereby synthesis errors can be decreased to arbitrarily small levels by improving quantisation accuracy sufficiently.

The third and final observation of this contribution is related to the time-evolution of the signal and perceptual models. Let $\mathbf{U}_k$, $\mathbf{L}_k$ and $\mathbf{H}_k$ correspond to the signal and perceptual models related to frame k. Conventionally, the models of frame k are applied strictly within the frame k, such that Eq. 4 becomes

$$\min_{\mathbf{x}_k} \|\mathbf{y}_k - \mathbf{U}_k\hat{\mathbf{x}}_{k-1} - \mathbf{L}_k\hat{\mathbf{x}}_k\|^2. \tag{12}$$

In contrast, for the modified objective function we define

$$\min_{\hat{\mathbf{x}}_k} \|\mathbf{H}_k(\mathbf{x}_k - \hat{\mathbf{x}}_k)\|^2 = \min_{\hat{\mathbf{x}}_k} \left\| \begin{bmatrix} \mathbf{L}_k \\ \mathbf{U}_k \end{bmatrix} (\mathbf{x}_k - \hat{\mathbf{x}}_k) \right\|^2 \tag{13}$$

Then, in the conventional model of Eq. 12, vector $\hat{\mathbf{x}}_k$ is multiplied with $\mathbf{L}_k$ in frame k and with $\mathbf{U}_{k+1}$ in frame $k+1$,

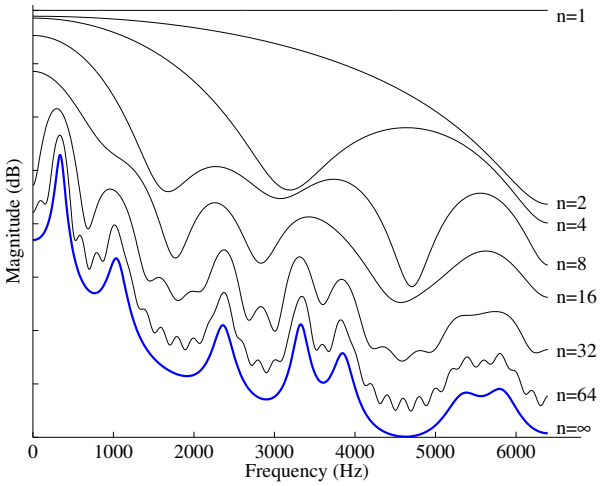

Figure 2: Effect of inconsistent signal and perceptual models demonstrated by truncating the impulse response of a typical vowel sound to different lengths n.

whereas in the modified objective function $\hat{\mathbf{x}}_k$ is multiplied with the model of only the same frame k, that is $\mathbf{H}_k$ (i.e. $\mathbf{U}_k$ and $\mathbf{L}_k$). In Figure 1 we then see that in the conventional approach, the impulse response applied to a sample x_n changes between positions h_{N-1-n} and h_{N-n}. In addition, in the objective function of Eq. 12, for the kth sample in the frame, only $N-k$ samples of the impulse response are taken into account. In other words, the impulse response is truncated to a length $N-k$. The effect of such truncations are illustrated in Fig. 2 with a typical voiced speech sound. Note that the spectra were shifted vertically for visual clarity.

In Fig. 2 we can then see that, as expected, when the impulse response is truncated to length $n=1$, the filter has a flat spectral envelope. With increasing length, an increasing amount of detail appears in the spectrum. However, with low lengths of the impulse response, such as $n=4$, spurious peaks can appear where none was in the original ($n=\infty$).

Clearly it cannot be a desired consequence that samples are perceptually weighted with a model whose accuracy depends on the position of the sample in a frame. Not only does the interpretation of the perceptual and signal models become blurred – which spectra of those illustrated in Fig. 2 correspond to our actual models? – but in addition, the performance of the overall model becomes hard to predict. From a scientific perspective it is then obvious that the modified objective function of Eq. 13 must be preferred over Eq. 12.

3. Application

To minimize perceptual error, we should minimize the norms in Eqs. 12 and 13. However, usually the quantized $\mathbf{x}_k$ is multiplied with a gain coefficient $\hat{\gamma}_k\hat{\mathbf{x}}_k$, whereby we can write

$$\|\mathbf{H}_k(\mathbf{x}_k-\hat{\gamma}_k\hat{\mathbf{x}}_k)\|^2=(\mathbf{x}_k-\hat{\gamma}_k\hat{\mathbf{x}}_k)^T\mathbf{R}_k(\mathbf{x}_k-\hat{\gamma}_k\hat{\mathbf{x}}_k). \quad (14)$$

The error energy then depends on two parameters, $\hat{\mathbf{x}}_k$ and $\hat{\gamma}_k$, which are dependent on each other. Joint minimization of $\hat{\mathbf{x}}_k$ and $\hat{\gamma}_k$ is thus difficult. To simplify the problem, the conventional approach is to begin by assuming that $\hat{\gamma}_k$ is optimal, then quantise $\mathbf{x}_k$ and finally determine $\hat{\gamma}_k$.

By calculating the optimal $\hat{\gamma}$ and inserting it into Eq. 14, we can readily obtain a new optimization problem which depends only on $\hat{\mathbf{x}}$:

$$\min_{\hat{\mathbf{x}}_k}\left[\frac{(\hat{\mathbf{x}}_k^T\mathbf{R}_k\hat{\mathbf{x}}_k)^2}{\hat{\mathbf{x}}_k^T\mathbf{R}_k\hat{\mathbf{x}}_k}\right]=\min_{\hat{\mathbf{x}}_k}\left[\frac{(\mathbf{d}_k'^T\hat{\mathbf{x}}_k)^2}{\hat{\mathbf{x}}_k^T\mathbf{R}_k\hat{\mathbf{x}}_k}\right] \quad (15)$$

where we defined the constant vector $\mathbf{d}_k'=\mathbf{R}_k\mathbf{x}_k$. Once the optimal quantisation $\hat{x}$ has been determined, the optimal gain can readily be calculated by

$$\hat{\gamma}_k=\frac{\mathbf{d}_k'\hat{\mathbf{x}}_k}{\hat{\mathbf{x}}_k^T\mathbf{R}_k\hat{\mathbf{x}}_k}. \quad (16)$$

Observe that since the model states that

$$\mathbf{y}_k=\mathbf{L}_k\mathbf{x}_k+\mathbf{U}_{k-1}\mathbf{x}_{k-1} \quad (17)$$

then $\mathbf{x}_n$ can be readily obtained by

$$\mathbf{x}_k=\mathbf{L}_k^{-1}(\mathbf{y}_k-\mathbf{U}_{k-1}\mathbf{x}_{k-1}). \quad (18)$$

Here $\mathbf{L}_k^{-1}$ can always be computed because $\mathbf{L}_k$ has full rank. Note, moreover, that Eq. 17 can also be used to synthesize the quantized signal $\hat{\mathbf{y}}_k$ using the quantized $\hat{\mathbf{x}}_k$'s.

As known from previous works [2], the objective function of the conventional form has similar equations

$$\min_{\hat{\mathbf{x}}_k}\left[\frac{(\mathbf{d}_k^T\hat{\mathbf{x}}_k)^2}{\hat{\mathbf{x}}_k^T\mathbf{L}_k^T\mathbf{L}_k\hat{\mathbf{x}}_k}\right] \quad \text{and} \quad \hat{\gamma}_k=\frac{\mathbf{d}_k\hat{\mathbf{x}}_k}{\hat{\mathbf{x}}_k^T\mathbf{L}_k^T\mathbf{L}_k\hat{\mathbf{x}}_k} \quad (19)$$

where

$$\mathbf{d}_k=\mathbf{L}_k^T\mathbf{L}_k\mathbf{x}_k \quad \text{and} \quad \mathbf{x}_k=\mathbf{L}_k^{-1}(\mathbf{y}_k-\mathbf{U}_k\mathbf{x}_{k-1}). \quad (20)$$

Since the two objective functions, Eqs. 15 and 19, differ only in the matrix of the denominator, the same iterative optimization algorithms can be applied in both cases.

4. Experiments

This work proposes a new objective function for ACELP which means that the objective measure of coding quality is changed. To be able to compare these two measures, we cannot use the measures themselves, since clearly a method optimized for measure X would win if the quality evaluation measure is X. With this challenge in mind, three experiments were designed, one with a stationary artificial signal, objective tests with POLQA using real speech data and subjective tests with expert listeners to verify the POLQA results.

As a first experiment, we thus tested a synthetic stationary signal with a known constant spectral envelope excited by a white-noise signal and encoded it using both the conventional error measures as well as the proposed error measure. The used spectral envelope is the same as in Figure 2 and represents a typical voiced sound. The sampling frequency was 12.8kHz and frame length 64 samples (corresponding to a sub-frame in AMR-WB). The pulse codebook consisted of 4 pulses which could be placed anywhere in the sub-frame (no track-structure was employed), the sign of the pulses were set to the same as the target residual and the positions of the pulses were determined with a pair-wise exhaustive search. As a lower reference, the same coding was conducted with the correlation matrix replaced by an identity matrix. All three codecs were implemented without LTP and other more advanced tools. For the gain of the pulse codebook we used Eq. 16 without quantisation.

The overall SNR of the codecs using conventional and proposed error measures, as well the lower reference, were respectively, 3.14dB, 3.22dB and 1.12dB. As expected, the proposed

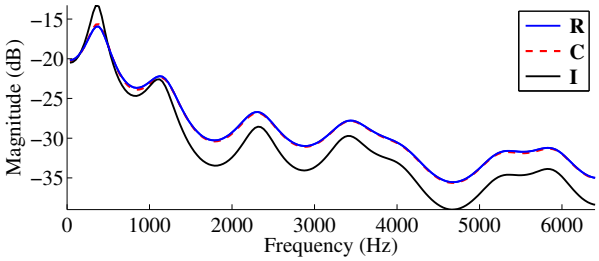

Figure 3: Illustration of the error spectra when coding using the autocorrelation matrix $\mathbf{R}$ in the proposed objective function Eq. 15, the correlation matrix $\mathbf{C}$ in Eq. 19, and the lower reference, where correlation matrix is replaced by the identity matrix $\mathbf{I}$ in Eq. 15.

error measure outperforms the conventional measure by 0.08dB and the lower reference is approximately 2dB worse. The improvement over the conventional method is modest, 4% relative to the lower-reference, but still a positive result. Informal experiments show that the result is highly dependent on the spectral shape, that spectral envelopes with sharp peaks yield much larger improvements and it is not difficult to create spectral envelopes where the improvement over the conventional approach is 0.5dB. Significant degradations were not observed for any spectral envelope shapes. Increasing the number of pulses also did not significantly change the results but severely increased the complexity of the brute-force pulse search.

The spectra of the error signals of the three simplified codecs are displayed in Fig. 3. Note that the flatter the error spectrum is, the more effective the optimization has been. We observe that the lower reference codec has a large dynamic range and that it follows the shape of the impulse response. In comparison, the two proper codecs have a flatter shape than the lower reference and the proposed error measure has a very slightly flatter shape than the conventional measure. Specifically, the dynamic ranges (ratio between highest and lowest peaks) for the conventional, proposed and lower reference measures are, respectively, 20.0dB, 19.6dB, and 25.9dB. This means that the proposed error measure follows the perceptual model 0.4dB more efficiently than the conventional measure.

The second experiment was to apply the new error measure in a full-scale ACELP implementation. We applied the consistent error measure into the pulse search, but not to gain or LTP calculations. The purpose of this choice was to obtain the maximal computational benefit, but minimize changes which would require re-tuning of perceptual parameters. This choice does, however, give the proposed method two disadvantages. Firstly, the proposed method will have a slight overhead in computational complexity, since having two different error measures requires that each of their own intermediate values have to be computed. Worse, since the algebraic codebook is optimized with the new objective function, then when optimizing the gains with the convetional objective function, the algebraic codebook will have an artificially lower correlation, and thus the gain will be attenuated. This compromise is necessary to get a reasonably fair comparison of methods, but has to be taken into account when analyzing results. Simultaneously, it demonstrates how inherently difficult it is to obtain fair comparisons of methods in this environment, where all tools have complex inter-dependencies.

The two objective functions were then tested for objective quality with POLQA [8]. The speech material used was the

Table 1: Objective comparison of conventional and proposed objective functions with confidence intervals.

Bit-rate (kbit/s)	Improvement in WMOPS	POLQA Conventional	Proposed
7.2	2.5%	3.50 ± 0.49	3.48 ± 0.50
8.0	14.5%	3.68 ± 0.48	3.66 ± 0.48
9.6	15,4%	3.81 ± 0.43	3.79 ± 0.44
13.2	17.3%	4.04 ± 0.39	4.04 ± 0.38

Table 2: The items with most extreme differences in POLQA scores and the corresponding differential MUSHRA scores (with confidence intervals) from a subjective listening test.

Item #	POLQA difference	MUSHRA difference
1	-0.22	0.89 ± 7.10
2	-0.22	0.44 ± 6.01
3	-0.20	-2.88 ± 3.23
4	0.17	-0.11 ± 6.28
5	0.21	-0.33 ± 6.63
6	1.03	0.00 ± 5.29

NTT AT Super Wideband Stereo Speech Database [9], which contains 6 languages. The results are listed in Table 1. At the lower bitrates, there is very slight degradation which is not, however, statistically significant. At 13.2 kbit/s, where also the computational advantage is the greatest, there is no degradation.

Both versions of the codec were instrumented to measure weighted millions operations per second (WMOPS) according to [10]. The relative improvement of the proposal in comparison to the conventional approach is shown in Table 1. The WMOPS values are calculated for the entire codec, not only the algebraic codebook, whereby in terms of the codebook optimization, the improvement is much larger.

Finally, 9 expert listeners verified the POLQA results by comparing the 3 items with the largest improvements and 3 items with the largest degradations. Table 2 lists the differential results of the MUSHRA test. Clearly, there is no statistically signficant difference in perceptual quality between the two codecs.

5. Conclusion

This contribution presents an improvement to the classical ACELP algorithm by applying the perceptual and signal models in a consistent manner, whereby computational complexity is reduced. If the number of iterations in the algebraic codebook optimization is kept constant, then the current implementation does not have a signficant impact on perceptual quality. However, several arguments show that the proposed method applies perceptual and signal models more consistently, that the modelling efficiency is improved for stationary signals and that the experiment structure put the proposed methods at an disadvantage. It is thus expected that a ground-up implementation based on the proposed approach could provide a perceptual improvement over the conventional method. By increasing the number of iterations to match the computational savings of the proposed methods, it is also possible to improve perceptual quality.

In conclusion, it was shown that the new error measure reduces coding noise for a synthetic stationary signal by 0.08dB and reduces the dynamic range of the error in the spectrum by 0.4dB. In addition, in an implementation to a full-scale codec, the proposed approach does not have a statistically significant impact on perceptual quality but reduces computational complexity by up to 17%.

6. References

[1] J.-P. Adoul, P. Mabilleau, M. Delprat, and S. Morissette, "Fast CELP coding based on algebraic codes," in *Acoustics, Speech, and Signal Processing, IEEE Int Conf (ICASSP'87)*, April 1987, pp. 1957–1960.

[2] B. Bessette, R. Salami, R. Lefebvre, M. Jelínek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, "The adaptive multirate wideband speech codec (AMR-WB)," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 8, pp. 620–636, 2002.

[3] ISO/IEC 23003-3:2012, "MPEG-D (MPEG audio technologies), Part 3: Unified speech and audio coding," 2012.

[4] F.-K. Chen and J.-F. Yang, "Maximum-take-precedence ACELP: a low complexity search method," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, vol. 2. IEEE, 2001, pp. 693–696.

[5] R. P. Kumar, "High computational performance in code exited linear prediction speech model using faster codebook search techniques," in *Proceedings of the International Conference on Computing: Theory and Applications*. IEEE Computer Society, 2007, pp. 458–462.

[6] N. K. Ha, "A fast search method of algebraic codebook by reordering search sequence," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, vol. 1. IEEE, 1999, pp. 21–24.

[7] M. A. Ramirez and M. Gerken, "Efficient algebraic multipulse search," in *Telecommunications Symposium, 1998. ITS'98 Proceedings. SBT/IEEE International*. IEEE, 1998, pp. 231–236.

[8] ITU-T Recommendation P.863, "Perceptual objective listening quality assessment," 2011.

[9] NTT AT, "Super wideband stereo speech database." [Online]. Available: http://www.ntt-at.com/products_e/widebandspeech/

[10] ITU-T Recommendation G.191, "Software tool library 2009 user's manual," 2009.